Template-based Monocular 3D Recovery of Elastic Shapes using Lagrangian Multipliers

Nazim Haouchine, Stéphane Cotin

Template-based Monocular 3D Recovery of Elastic Shapes using Lagrangian Multipliers

Nazim Haouchine and Stephane Cotin
Inria - Mimesis Group
1, place de l'Hopital, Strasbourg, France
nazim.haouchine@inria.fr | stephane.cotin@inria.fr

Abstract

We present in this paper an efficient template-based method for 3D recovery of elastic shapes from a fixed monocular camera. By exploiting the object's elasticity, in contrast to isometric methods that use inextensibility constraints, a large range of deformations can be handled. Our method is expressed as a saddle point problem using Lagrangian multipliers resulting in a linear system which unifies both mechanical and optical constraints and integrates Dirichlet boundary conditions, whether they are fixed or free. We experimentally show that no prior knowledge on material properties is needed, which exhibit the generic usability of our method with elastic and inelastic objects with different kinds of materials. Comparisons with existing techniques are conducted on synthetic and real elastic objects with strains ranging from 25% to 130% resulting to low errors.

1. Introduction

Three-dimensional reconstruction of non-rigid objects from a monocular camera is a complex problem with many potential applications in computer graphics, augmented reality and surgical vision. The complexity emanates from its underconstrained nature since several shape configurations may produce the same projection which leads to ambiguities [6]. To resolve these ambiguities considering inextensible materials, various approaches have been considered that, most of the time, rely on additional constraints such as the preservation of mesh geometrical properties, temporal consistency or shading information. However, the additional constraints that are used are not always suited to the properties of an elastic object.

Physics-based models have been seldom used in the past due the complexity which precludes real time algorithms and due to the necessary assumption on the material stiffness (which is difficult to acquire). Nevertheless, recent studies [2, 13, 8, 12, 3] have shown that such methods are appropriate when dealing with elastic objects that can undergo stretching or compression. Thus, we advocate in this paper the use of elastic models to enable reconstruction of deformable objects undergoing large deformations (see Figure 1). However, in contrast to related works that express the problem as force minimization [8] [12] or as non-linear energy minimization [13], we propose to express the problem as a saddle point problem using Lagrangian multipliers. This formulation permits to build a well-posed linear system that unifies both optical and mechanical constraints. Our method is invariant of material properties which makes it adequate to elastic and inelastic deformations, and behave well with poorly textured surfaces and presence of occlusions.

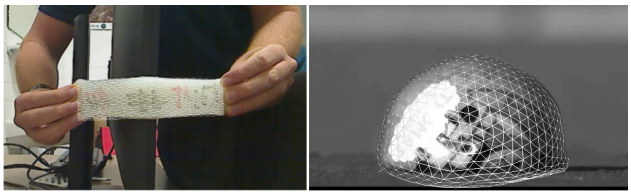

Figure 1: 3D recovery of elastic objects from a fixed monocular camera. Instead of minimizing the forces that act on the object, our method solves a constrained optimization that permits to handle elastic and inelastic materials without a prior on their properties.

We therefore view the contribution of this paper as a generic method that: *(i)* truly handles large elastic deformations up to 130 % (more than 2 times the initial shape), *(ii)* is invariant to material properties and get rid of mass, damping coefficients, time-steps and external forces, which are one of the major concerns in using mechanical models in computer vision tasks, and *(iii)* is adapted to both extensible and inextensible materials as well as textured and poorly textured surfaces and behave well with the presence of occlusions. Furthermore, we experimentally validate it with various type of materials and data, with tests on synthetic data and real data with quantitative measures and quantitative assessment and comparison with related works.

2. Related Works

Recovering 3D shapes from a monocular camera has been an extensive area of research in the last decade. Prior works have first considered inextensible surfaces where the exploitation of the underlying distance constraints is possible, using parametric geometrical warping models [18] [22] [17], a set of representative sample of possible shapes from learning-based approaches [21] [19], or by considering the problem as a convex minimization of the reprojection error [20] [23].

Previous methods are very effective when considering inextensible surfaces. However, they are not suitable for elastic objects, where the conservation of shape's geometrical properties cannot be considered as a plausible constraint. One way to overcome this issue is to rely on environment light and shading information. To this end, a closed-form method was introduced in [14] to capture stretching surfaces. This method assumes a Lambertian surface with a single point light source and yields good results. However, the strong assumption on the lighting makes the method hard to generalize in all environments.

The closest works to our method are the ones that deal with elastic shapes using mechanical models [13, 8, 12, 9] The approach described in [13] relies on the minimization of a stretching energy subject to external image constraints. The problem is formalized as a non-linear minimization and uses an underlying B-spline model as a regularization term. It unifies geometric and mechanical constraints assuming local linear elasticity and shows effective results considering Poisson's ratio as the unique mechanical parameter. This method was improved in [12] by integrating fixed boundary conditions. Nevertheless, it was not particularly designed to handle large deformations where the results shown are limited to an extensibility of 15%. In a similar context, non-linear elasticity has been considered in [8], for the reconstruction of highly elastic objects. The problem was expressed as force minimization problem subject to fixed boundary conditions and considering the image points as stretching forces. The method provided convincing results and was later extended to handle self-occlusions [9]. However, it requires the definition of internal stretching and external stretching parameters, which are hard to estimate in general. In [2], the authors combined a physical model derived from Navier's equations with an Extended Kalman Filter to efficiently estimate 3D elastic shapes while simultaneously compute the camera pose. Similar to this, [3] proposed to incorporate a dynamic particle model into a bundle adjustment framework. Both methods exhibit convincing results under small elastic deformations.

3. Elastic Model

Our approach focuses on recovering the 3D deformation of an object from its 2D projection on the image. It relies on the use of an elastic model that describe the behavior of the recovered surface. In this section, we will describe the material law that defines the elastic model, the discretization of the model using Finite Element Method (FEM) and the resolution of the system through a static integration scheme.

3.1. Constitutive Law

The literature related to deformable models is vast and crosses many scientific domains such as engineering, computational mechanics, computer graphics [15]. Several characteristics are sought for our elastic model: a low computation cost while having reasonable accuracy, the ability to handle large deformations and a low parameterization. While dedicated models such as thin-plate splines or free-from have proved their relevance and efficiency for isometric deformations such ones that undergo paper sheets, sails or cloths [5], we intend to have a more generic model in order to handle elastic materials such as silicone samples, hair ribbon or liver tissue. For this purpose, a Saint-Venant Kirchoff material appears to be a relevant strategy [8]: the material is hyper-elastic, thus allowing to handle large deformations, it relies on few material parameters and can be quickly computed thanks to the work in [10].

A Saint-Venant Kirchoff material is defined by a stress-strain relationship of the form:

$$\mathbf{S} = \eta(tr\mathbf{E})\mathbf{I}_3 + 2\mu\mathbf{E} \tag{1}$$

where $\mathbf{S}$ is the second Piola stress tensor, $\mathbf{E}$ is the Green-Lagrange strain tensor and $tr\mathbf{E}$ its trace, $\mathbf{I}_3$ is the 3×3 identity matrix and η and μ are Lamé coefficients and can be computed thanks to the elastic parameters of the material E and ν. E is Young's modulus and is a measure of the stiffness of the material while ν is Poisson's ratio and estimates the compressibility of the material.

3.2. Discretization with Finite Element Method

Without loss of generality, we use the FEM to discretize the partial differential equations in (1). The deformable object is represented as a volumetric mesh consisting of 3D polyhedra called elements. We choose here to rely on therahedral elements with linear shape functions [10].

A particular object deformation is specified by the displacements of mesh vertices (nodal positions) and/or the nodal forces. In general, the relationship between nodal forces and nodal positions is nonlinear. When linearized, the relationship for an element e connecting n_e nodes can simply be expressed as $\mathbf{f_e} = \mathbf{K_e}\mathbf{u_e}$, where $\mathbf{f_e} \in \mathbb{R}^{3n_e}$ contains the n_e nodal forces and $\mathbf{u}_e \in \mathbb{R}^{3n}$ the n_e nodal displacements of an element. The matrix $\mathbf{K_e} \in \mathbb{R}^{3n_e \times 3n_e}$ is

called the stiffness matrix of the element. Because elastic forces coming from adjacent elements add up at a node, the stiffness matrix $\mathbf{K} \in \mathbb{R}^{3n\times 3n}$ for an entire mesh with n nodes is formed by assembling the element's stiffness matrices $\mathbf{K_e}$. The equation of a deformation on an elastic object will therefore take the form:

$$\mathbf{K}\mathbf{u} = \mathbf{f} \tag{2}$$

The computation of the stiffness matrix $\mathbf{K}$ is nonlinear due to the non-linearity of Green-Lagrange strain tensor and should be recomputed after every deformation to remain valid. Because we want to handle several kinds of elastic deformations, the stiffness matrix is computed as follow:

$$\mathbf{K}(\mathbf{u}) = \frac{\partial \mathbf{f}(\mathbf{u})}{\partial \mathbf{u}} \tag{3}$$

thus allowing for a large displacement $\mathbf{du}$ (such as $\mathbf{u}' = \mathbf{u} + d\mathbf{u}$) of the vertices of the mesh.

We can notice that the deformation depends only on the computation of internal forces. Indeed, the solution does not integrate acceleration or velocity, since we choose here no to consider kinetic energy. We address here a static scenario where only a rest and a deformed image states are provided. This reduces the number of parameters involved in the solution (mass, damping coefficients, time-step) that are usually considered in dynamic scenario (time-stepping techniques [4]).

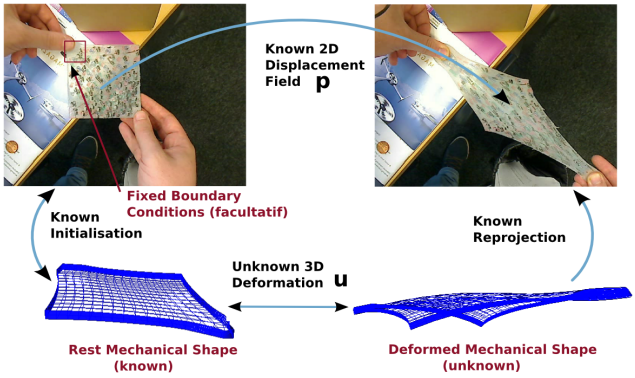

Figure 2: Problem formulation: we aim at recovering the elastic deformed shape corresponding to $\mathbf{u}$, from the known reprojected displacement in the image $\mathbf{p}$, using a physical model at rest shape and possible fixed boundary conditions.

4. Problem formulation

We follow the formalization of [8] and [12] that aims at unifiying the physical constraints of the targeted object, represented by the internal forces that emanate from its constitutive law, and its optical protective constraints that represent the projection of the elastic object on the image with respect to the camera parameters (see Figure 5).

Assume a physical model with n nodes and assuming m feature correspondences between the deformed and the rest shape of the object in the image. Let $\mathbf{u}^{rest} \in \mathbb{R}^{3n}$ be the positions of our physical model at its rest configuration and $\mathbf{u}^{def} \in \mathbb{R}^{3n}$ the unknown position at deformed configuration. And let $\mathbf{p}^{rest} \in \mathbb{R}^{2m}$, be the vector of extracted features at rest configuration and $\mathbf{p}^{def} \in \mathbb{R}^{2m}$ their position at the deformed state.

Given a projection matrix $\mathbf{P}$, the optical constraints that relate each 3D point u^i in world coordinates to the 2D point p_i in image coordinates are expressed as follows

$$\mathbf{P}u_i = p_i, \qquad \text{for } i = 1, \dots m \tag{4}$$

Considering the mechanical and optical constraints, and by letting $\mathbf{u} = \mathbf{u}^{rest} - \mathbf{u}^{def}$ and $\mathbf{p} = \mathbf{p}^{rest} - \mathbf{p}^{def}$. The problem can be formalized as finding $(\mathbf{u}^{rest} + \mathbf{u}) \in \mathbb{R}^{3n}$ such that:

$$\begin{cases} \mathbf{K}\mathbf{u} &= \mathbf{f} \\ \mathbf{L}\mathbf{u} &= \mathbf{p} \end{cases} \tag{5}$$

where $\mathbf{L} \in \mathbb{R}^{2m\times 3n}$ is built from the projection matrix $\mathbf{P}$.

5. Elastic Shape Recovery

5.1. Saddle Point Problem

We aim at recovering $(\mathbf{u}^{rest}+\mathbf{u})$ that it satisfies both material and image constraints. Where related works proposed to solve the set of equations 5 by minimizing the external forces $\mathbf{f}$. We express our problem as a saddle point problem resulting in a linear system that can be solved with iterative of direct solvers. To do so, we use Lagrangian multipliers method that aims at finding the local minima of a function subject to equality constraints, as follows.

The potential energy of the unconstrained finite element model can take the form

$$\mathcal{W} = \frac{1}{2}\mathbf{u}^\top\mathbf{K}\mathbf{u} - \mathbf{u}^\top\mathbf{f} \tag{6}$$

Using the potential energy $\mathcal{W}$, we can express the system of equations 5 as the following minimization problem:

$$\min_{\mathbf{u}\in\mathbb{R}^{3n}} \{\frac{1}{2}\mathbf{u}^\top\mathbf{K}\mathbf{u} - \mathbf{u}^\top\mathbf{f} : \mathbf{L}\mathbf{u} = \mathbf{p}\} \tag{7}$$

We can turn this constrained quadratic minimization into an unconstrained problem using the theory of Lagrangian multipliers by optimizing the Lagrangian function:

$$\mathcal{L}(\mathbf{u}, \boldsymbol{\lambda}) = \underbrace{\frac{1}{2}\mathbf{u}^\top\mathbf{K}\mathbf{u} - \mathbf{u}^\top\mathbf{f}}_{\text{Elastic constraints}} + \underbrace{\boldsymbol{\lambda}^\top(\mathbf{L}\mathbf{u} - \mathbf{p})}_{\text{Projective constraints}} \tag{8}$$

Here, we adjoined $2m$ Lagrangian multipliers collected in vector $\boldsymbol{\lambda}$. Extremizing $\mathcal{L}$ with respect to $\mathbf{u}$ and $\boldsymbol{\lambda}$ yields the multiplier-augmented form

$$\begin{bmatrix} \mathbf{K} & \mathbf{L}^\top \\ \mathbf{L} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{p} \end{bmatrix} \tag{9}$$

The stiffness matrix $\mathbf{K}$ is bordered with $\mathbf{L}$ and $\mathbf{L}^\top$, where the vector $\mathbf{p}$ contains the boundary conditions extracted from the image, while the vector $\boldsymbol{\lambda}$ can be interpreted as forces required to maintain the boundary conditions $\mathbf{p}$. Finally, the solution $\mathbf{u}$ is the vector of saddle points of the problem and its uniqueness is discussed below.

5.2. Compliance matrix relaxation

Ideally, we would want the reprojection error to be zero for all u_i for which we have a reprojected image point p_i. However, in practice, due to noisy image measurements, this is never possible. The formulation of Eq. 10 tend to exactly satisfy the reprojection constraints, which can be highly damaging in the presence of outliers. Therefore, we introduce an additional variable to relax the reprojection constraints and rewrite our problem as

$$\begin{bmatrix} \mathbf{K} & \mathbf{L}^\top \\ \mathbf{L} & \mathbf{C} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{p} \end{bmatrix} \tag{10}$$

where $\mathbf{C} \in \mathbb{R}^{2m \times 2m}$ can be seen as a compliance matrix which is composed of an uncertainty σ_i associated to each feature p_i. Following the argumentation in [24] we choose the inverse of the Hessian as the covariance for the measure for feature localization uncertainty.

6. Implementation

6.1. Initialization of forces vector

The question of initialization of the force vector $\mathbf{f}$ naturally arises from the system of equation 10. At rest configuration the vector of external forces $\mathbf{f}$ is a null vector, leading to solve Eq. 10 where both $\mathbf{K}$ and $\mathbf{u}$ are known. Moreover, forces controlling the deformation of the object, and derived from the measured displacement, are automatically computed from the Lagrange multipliers and do not need to be set.

6.2. Building the L matrix

In order to build the matrix $\mathbf{L}$, we first use the intrinsic matrix to represent the features in world units. Then, assuming an orthographic camera, the projection of the 3D points $\mathbf{u}$ onto the image can be expressed as:

$$p_i = \mathbf{R}u_i + \mathbf{T}, \qquad \text{for } i = 1, \dots m \tag{11}$$

where $\mathbf{R}$ is a 2×3 Stiefel matrix that encodes the two first rows of the camera rotation matrix and $\mathbf{T}$ is a 2×1 translation vector of the form

$$\mathbf{R} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \end{pmatrix}, \qquad \mathbf{T} = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \tag{12}$$

Assuming the features are registered to the centroid of the object and considering that $(n > m)$ (to avoid over-constrained system), the sparse matrix $\mathbf{L}$ of size $2m \times 3n$ is built from the rotation matrix $\mathbf{R}$ so that

$$\mathbf{L}_n = \begin{cases} \mathbf{R}, & \text{if } n \in m \\ \mathbf{0}_{(2,3)}, & \text{otherwise} \end{cases} \tag{13}$$

6.3. Mapping features with nodal positions

In practice, image points do not coincide with the nodes of the mechanical model. At rest configuration, we can express each feature p_i using barycentric coordinates[1] of facet vertices, such that $p_i^{rest} = \sum_{j=1}^{3} \phi_j(x_i, y_i) u_j^{rest}$, where $\phi(x, y) = a + bx + cy$ with (a, b, c) being the barycentric coordinates of the triangle composed of nodal points u_j^{rest}, with $1 \leq j \leq 3$. This linear relation remains valid during the deformation which permits to express features positions as a linear combination of the mechanical nodal positions.

6.4. Boundary conditions

We refer in this study to boundary conditions as the *Dirichlet boundary conditions* that can be seen as a set of displacements that is imposed on the mechanical nodes. These boundary conditions can be fixed (the displacement is null) and are called homogeneous, or can follow a prescribed displacement that usually emanates from external forces such as gravity, elongation, torsion or compression and are called heterogeneous. In our formulation, this set of boundary conditions represents features and forms the sets $\{\mathbf{p}^{rest}, \mathbf{p}^{def}\}$ and they implicitly describe both homogeneous boundary conditions and heterogeneous boundary conditions. Mathematically speaking, denoting $\mathbb{L}$ the domain of the volume mesh and $\mathbb{S} = \partial\mathbb{L}$ its boundary (i. e. the surface), we define:

- Homogeneous boundary conditions: $u_i = 0$ for $i \in \mathbb{L}_f$ where $\mathbb{L}_f \subset \mathbb{L}$ is the part of the surface which is fixed, i. e., the nodes do not move during the resolution.
- Heterogeneous boundary conditions: $u_i = p_i$ for $i \in \mathbb{L}_m$ where $\mathbb{L}_m \subset \mathbb{L}$ and $\mathbb{L}_m \cap \mathbb{L}_f = \emptyset$ are prescribed positions.

The presence of homogeneous boundary conditions is not necessary for solving the system of equation 10. In practice, their identification in the scene brings useful information for the modeling of the object and can improve 3D recovery in certain cases, as it will be shown in Section 7.

7. Results

We present in this section the results obtained using our method and the comparison conducted with existing techniques. We report results obtained on real and synthetic

[1] Equivalent to using linear shape functions defined on P1 tetrahedral elements.

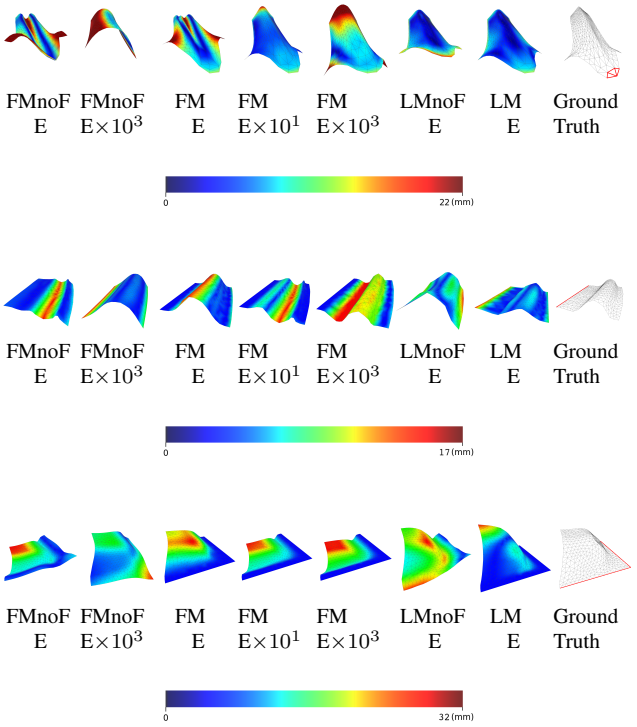

Figure 3: Results on synthetic data with variation of Young's Modulus E in the range $E \times 10^{-1}$ to $E \times 10^{4}$, with and without fixed boundary conditions (shown in red in the ground truth). [better seen in color]

data to illustrate the ability of our approach to capture 3D large elastic deformations. We first test our approach on synthetic data to show the invariance of our formulation to elastic parameters and its robustness to noise and lack of features. We then use the dataset of [8] where several video sequences of a silicone-made object undergoing different types of stretching deformation is proposed. We quantify the three-dimensional shape recovery error with respect to a ground truth. We also conduct experiments on inextensible surfaces and on poorly textured objects to highlight the potential uses of our method. In all experiments we used SIFT [11] to detect 2D features.

We compare our Largangian multipliers method denoted **LM** with existing approaches: an inextensible approach that handles isometric deformations described in [5] that we denote **Inext**, a physically-based approach for elastic surfaces that formulate the problem as force minimization described in [8] that we denote **FM** and linear least-square solution that uses linear elasticity with fixed boundary conditions encoded directly in the stiffness matrix [12] that we denote **LLS**. We also consider the impact of fixed boundary conditions in our experiments with the methods **FMnoF** and **LMnoF** that respectively denote the **FM** and **LM** methods without fixed boundary conditions. In addition to quantitative comparisons with [20] a convex optimization method and [16] denoted **CVX** a recent approach based on Laplacian meshes denoted **LAP**, both dedicated to inextensible 3D shape recovery.

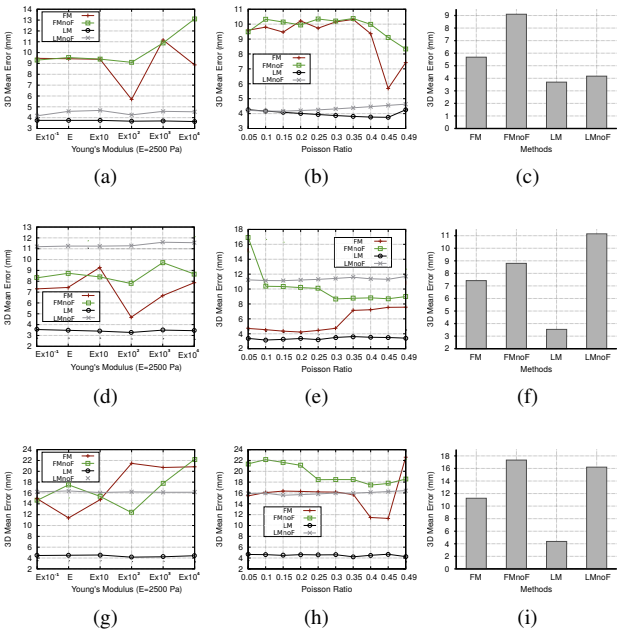

Figure 4: Results on synthetic data: (a)(d)(g) sensitivity to Young's Modulus E; (b)(e)(h) sensitivity to Poisson's ratio ν; (c)(f)(i) 3D mean errors for each method: our method exhibits the lowest error while being quasi-invariant to material properties.

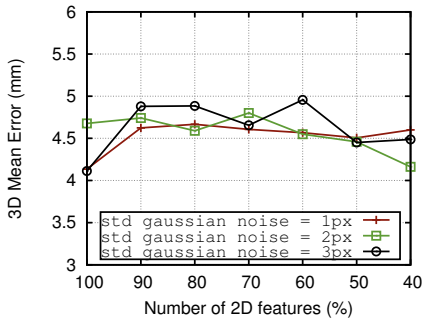

Figure 5: Results on synthetic data: Robustness to image noise and lack of features: randomly reducing the number of constraints (from 144 features to 58 features) and disturbing the data with Gaussian noise weakly impact 3D reconstruction error in a 1 mm interval.

Method	Inext	LLS	FM	LM	Inext	LLS	FM	LM
Exp.	Mean				RMS			
Def1	3.62	3.28	1.99	**1.66**	4.29	3.82	2.62	**1.93**
Def2	2.77	1.94	2.00	**1.27**	3.47	2.45	2.44	**1.60**
Def3	2.98	2.06	1.56	**1.55**	3.71	2.58	2.06	**1.91**
Def4	4.87	1.99	**1.19**	1.31	5.58	2.53	1.69	**1.51**

Table 1: Comparison with related works for each deformation with mean 3D error (in mm) and RMS 3D error (in mm).

7.1. Synthetic Data

We synthesize elastic deformations on a simulated silicone-made object of size $100 \times 100 \times 10$ mm^3. The object is composed of 432 linear tetrahedral elements characterized by a Young's Modulus $E = 2500$ Pa and a Poisson's ratio $\nu = 0.45$. We apply forces on the simulated object to produce deformed shapes with a plan elongation in a range between 40% and 120% and a depth elastic deformation between 30% to 50%. Video sequences of 640×480 images are acquired using a virtual camera with focal length $f_u = f_v = 500$ and principal point at $(u_c, v_c) = (320, 240)$. On this dataset we run the following experiments: ***(1) sensitivity to elastic paramaters:*** we produce elastic deformations for each of the abovementioned method by varying the value of a Young's Modulus E in the range $E \times 10^{-1}$ to $E \times 10^4$ and a Poisson's ratio ν from 0.05 to 0.49 [2] while keeping the same amount of boundary conditions. ***(2) robustness to image noise and lack of features:*** a Gaussian noise with standard deviation $g_{std} \in \{1px, 2px, 3px\}$ is added to the features. In addition, the number of these features was randomly reduced to reach 40% of the initial number (144 features).

For each set we compute a 3D mean error (in mm) as the vertex-to-vertex distance between the reconstructed mesh and the ground-truth mesh. The corresponding plots are reported in Figures 4 and the resulting shapes with error measurements are illustrated in Figure 3.

Our **LM** method produce the lowest errors w.r.t other methods. We can notice that the recovery error with **LM** and **LMnoF** are quasi-independent to Young's modulus and Poisson's ratio values, while their variation strongly impacts **FM** and **FMnoF** techniques. The results also show that, depending on the simulation, fixing a subset of the boundary conditions highly reduces the errors. Which suggest that a correct placement of fixed boundary conditions can be as important as the physical modeling. This difference is even more important when the fixed nodes represents a large part of the object.

We can also notice in figure 5 that our method works well with a reduced number of boundary conditions perturbed with noise, where the impact of these perturbations on the 3D mean error is less than 1 mm. This demonstrates that the **LM** approach is appropriate to poorly textured surfaces as it will be shown with real data.

[2] Note that the values $\nu = 0$ and $\nu = 0.5$ are excluded since the lame coefficient $\eta = E\nu/(1+\nu)(1-2\nu)$

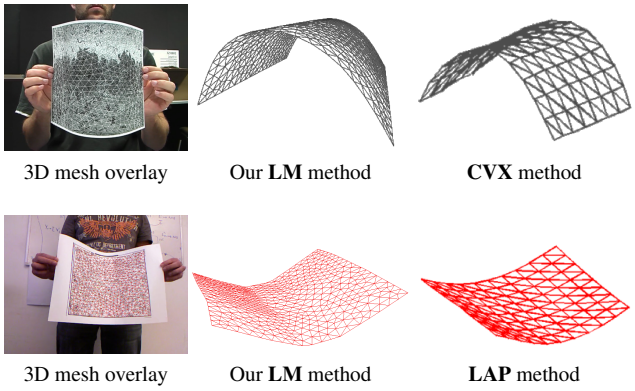

Figure 10: Results on inelastic data with smoothly deforming papers with rich textures. [better seen in color]

7.2. Real Data with Ground Truth

We test our approach with real data from the silicone-elastic dataset [8]. This dataset consists of a silicone-made object deformed following several configurations with extensibility ranging from 25% to 120%. The silicone strip has a size of $100 \times 100 \times 10$ mm^3 and its stiffness is characterized by a Young's Modulus $E = 250000$ Pa and a Poisson's ratio $\nu = 0.45$. For each configuration a video sequence with image resolution of 640×480 is acquired with a monocular camera at 30 fps. We only exploit the first and last frames of this video sequence. The 3D shape at final state is provided and is considered as ground truth. The resulting 3D shapes are illustrated in Figures 6, 7, 8, 9 and the comparison is reported in 1.

The **LM** method gives the lowest errors (Deformation 1, Deformation 2 and Deformation 4), or very close to the **FM** method (Deformation 3), with errors below 2 mm. In general, **FM** method gives relatively good results. **Inext** method that was designed for isometric deformations, fails at recovering the 3D shape especially with Deformation 4 where large elastic deformations occur. The **LLS** method gives good results but limited to linear elasticity. The integration of fixed boundary conditions reduces the error for both **FM** and **LM**. The **LM** approach is used without specific material properties while the **FM** approach requires to set the stiffness of external forces represented as springs and material properties represented by the stiffness matrix to reach an equilibrium between internal and external forces. Overall, the execution time varies between 0.2 seconds and 1.7 seconds, depending on the size of the system (mesh resolution and number of features).

Figure 6: Deformation 1. A folding deformation is produced with an elasticity of 30%.

Figure 7: Deformation 2. The silicone strip is constrained by a beam which produced a deformation with elasticity of 25%.

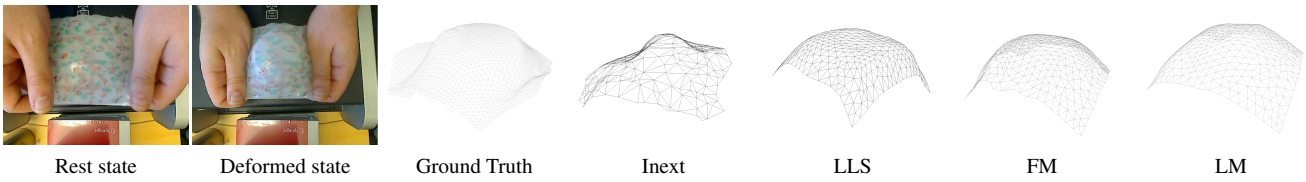

Figure 8: Deformation 3. Constrained silicone strip by a circular obstacle leading in a deformation of 40% of elasticity.

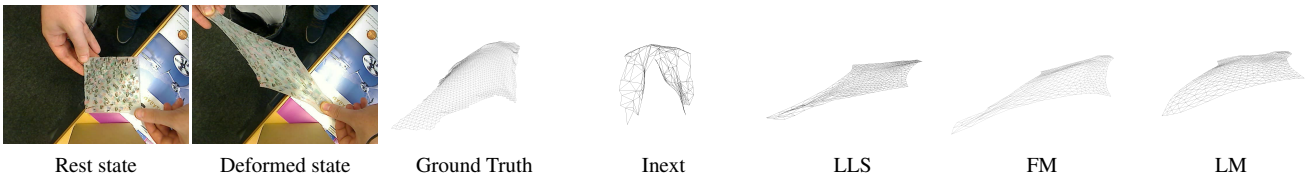

Figure 9: Deformation 4. A large elongation is produced with an elasticity of 130%.

7.3. Real Data without Ground Truth

Inextensible Paper Bending Experiments on synthetic data exhibit the invariance of our method to material properties. In order to confront our method to a real scenario, we used the data of [21] and [16] representing a smoothly deforming sheet of paper with rich texture. We also visually compare with these two methods. The results illustrated in figure 10 show that our method performs well without any knowledge of material properties.

Poorly Textured and Occluded Silicone Strip We further confront our method to a low textured object to highlight its performance in this kind of scenarios (cf figure 11). The experiment involves a soft object of $180 \times 30 \times 20$ mm^3 that is simulated with 240 linear tetrahedral elements and unknown material properties. The object is constrained by a rigid obstacle and elongated in-depth in several directions. From the acquired images we used only 22 SIFT features as boundary conditions with the full example and 16 SIFT features with the occluded one. The results exhibits correct 3D shape recovery in both cases (c.f figure 11).

Colliding Soft Ball Here we present our results on a soft ball colliding the ground in slow motion (youtube video). As for the other tests, no prior knowledge of material properties in considered. The spherical volume model is composed of 512 linear P1 tetrahedral elements. The recovery and augmentation is performed in real-time at 25 fps and is illustrated in figure 12.

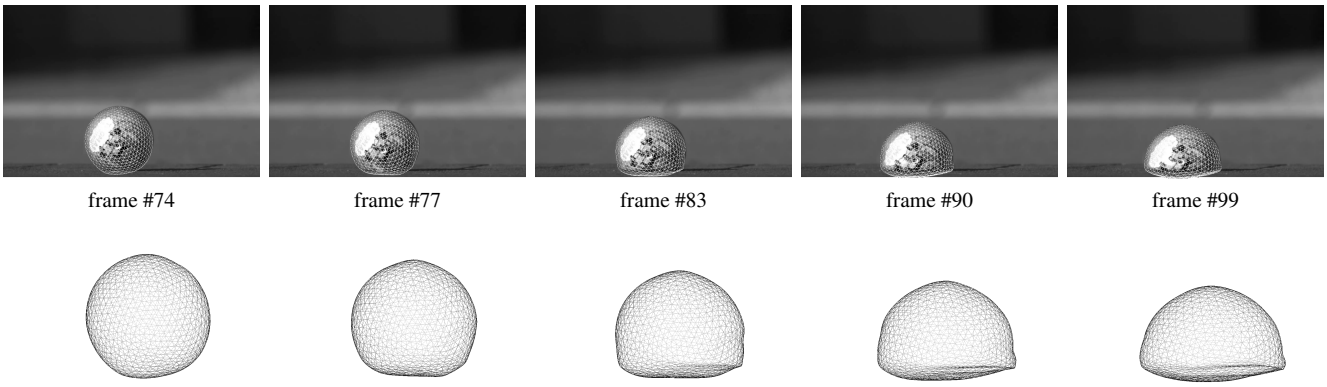

Figure 12: Rubber ball colliding the ground in slow motion.

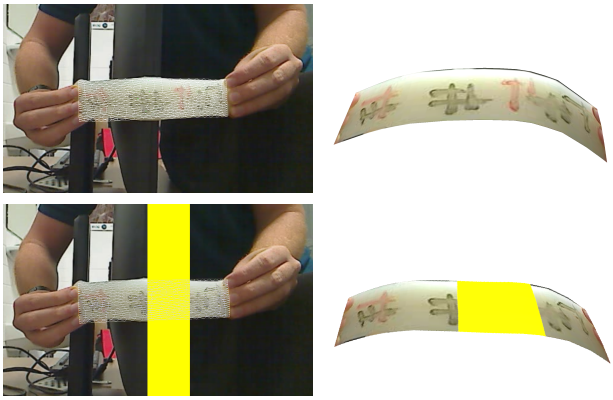

Figure 11: Results on poorly textured elastic data with a soft object (left) meshes overlay on the images and (right) the recovered 3D shape with retexturing. [better seen in color]

8. Discussion

Thanks to the use of a quasi-static integration scheme, the number of physical parameters has been reduced. Although it can be sufficient for many cases, it shows limits when transient behaviour like oscillations or vibrations have to be captured (the soft colliding ball in slow motion is a good example). To this end, a dynamic integration scheme can be used with the drawback of additional parameters, object's mass and damping coefficient for instance.

The major limitation of our method are the possible mismatched points between the rest and deformed configuration. Although we clearly make the assumption that we have m features correspondences to formalize the problem, in practice such assumption is hard to ensure. Yet, to the best of our knowledge, no descriptor is robust enough to handle large elastic deformations, since these deformations produce large geometric changes and texture variance on objects. One possible solution will be to use a learning approach and rely on the physical model to pre-compute shapes and texture to feed the learning model.

Moreover, relaxing the assumption of fixed camera is an important issue to be addressed. This implies to decompose rigid and non-rigid motion to compute deformations while estimating camera pose. To this end, one can pre-compute modal shapes using modal analysis [1] or decompose the stress tensor to extract rigid components [7].

9. Conclusion

We studied in this paper the problem of 3D recovery of elastic surfaces from a fixed monocular camera using mechanical models. We present a generic method that is robust to the choice of material properties and performs well on low textured surfaces. Our formulation unifies both mechanical and optical constraints by considering the problem as a saddle point problem, and integrated boundary conditions in en elagant manner. This results in a linear system that can efficiently be solved by adjoining Lagrangian multipliers. We conducted several experiments showing convincing results on elastic and inelastic deformations, and comparisons with state-of-the art techniques. Finally, moving scope techniques that simultaneously estimate pose and 3D shape of elastic objects can benefits from the output of this paper.

References

[1] A. Agudo, L. Agapito, B. Calvo, and J. M. M. Montiel. Good vibrations: A modal analysis approach for sequential non-rigid structure from motion. In *Com-*

puter Vision and Pattern Recognition (CVPR), 2014. IEEE Conference on, pages 1558–1565, 2014. 8

[2] A. Agudo, B. Calvo, and J. M. M. Montiel. Finite element based sequential bayesian non-rigid structure from motion. In *Computer Vision and Pattern Recognition, 2012. CVPR'12. IEEE Conference on*, pages 1418–1425, 2012. 1, 2

[3] A. Agudo and F. Moreno-Noguer. Simultaneous pose and non-rigid shape with particle dynamics. In *Computer Vision and Pattern Recognition (CVPR), 2015. IEEE Conference on*, pages 2179–2187, 2015. 1, 2

[4] M. Anitescu, F. A. Potra, and D. E. Stewart. Time-stepping for three-dimensional rigid body dynamics. *Computer methods in applied mechanics and engineering*, 177(3):183–197, 1999. 3

[5] A. Bartoli, Y. Gerard, F. Chadebecq, and T. Collins. On template-based reconstruction from a single view: Analytical solutions and proofs of well-posedness for developable, isometric and conformal surfaces. In *Computer Vision and Pattern Recognition (CVPR), 2012 Conference on*, pages 2026–2033, 2012. 2, 5

[6] F. Brunet, A. Bartoli, and R. Hartley. Monocular template-based 3d surface reconstruction: Convex inextensible and nonconvex isometric methods. *Computer Vision and Image Understanding*, 125(0):138 – 154, 2014. 1

[7] N. Haouchine, M.-O. Berger, and S. Cotin. Simultaneous pose estimation and augmentation of elastic surfaces from a moving monocular camera. In *ISMAR*, pages 199–202, 2016. 8

[8] N. Haouchine, J. Dequidt, M.-O. Berger, and S. Cotin. Single view augmentation of elastic objects. In *ISMAR*, pages 199–208, 2014. 1, 2, 3, 5, 6

[9] N. Haouchine, J. Dequidt, M.-O. Berger, and S. Cotin. Monocular 3d reconstruction and augmentation of elastic surfaces with self-occlusion handling. *IEEE Transactions on Visualization and Computer Graphics*, 21(12):1363–1376, 2015. 2

[10] R. Kikuuwe, H. Tabuchi, and M. Yamamoto. An edge-based computationally efficient formulation of saint venant-kirchhoff tetrahedral finite elements. *ACM Transactions on Graphics (TOG)*, 28(1):8, 2009. 2

[11] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, Nov. 2004. 5

[12] A. Malti, A. Bartoli, and R. Hartley. A linear least-squares solution to elastic shape-from-template. In *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*, pages 1629–1637, June 2015. 1, 2, 3, 5

[13] A. Malti, R. Hartley, A. Bartoli, and J.-H. Kim. Monocular template-based 3d reconstruction of extensible surfaces with local linear elasticity. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 1522–1529, 2013. 1, 2

[14] F. Moreno-Noguer, M. Salzmann, V. Lepetit, and P. Fua. Capturing 3d stretchable surfaces from single images in closed form. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1842–1849, June 2009. 2

[15] A. Nealen, M. Müller, R. Keiser, E. Boxerman, and M. Carlson. Physically based deformable models in computer graphics. In *Computer Graphics Forum*, volume 25 (4), pages 809–836, 2006. 2

[16] J. Östlund, A. Varol, D. T. Ngo, and P. Fua. Laplacian meshes for monocular 3d shape recovery. In *European Conference on Computer Vision (ECCV) 2012*, pages 412–425, 2012. 5, 7

[17] M. Perriollat, R. Hartley, and A. Bartoli. Monocular template-based reconstruction of inextensible surfaces. *International Journal of Computer Vision*, 95(2):124–137, 2011. 2

[18] D. Pizarro and A. Bartoli. Feature-based deformable surface detection with self-occlusion reasoning. *International Journal of Computer Vision*, 97(1):54–70, 2012. 2

[19] M. Salzmann and P. Fua. Linear local models for monocular reconstruction of deformable surfaces. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(5):931–944, May 2011. 2

[20] M. Salzmann, R. Hartley, and P. Fua. Convex optimization for deformable surface 3-d tracking. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8, Oct 2007. 2, 5

[21] M. Salzmann, J. Pilet, S. Ilic, and P. Fua. Surface deformation models for nonrigid 3d shape recovery. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(8):1481–1487, Aug. 2007. 2, 7

[22] A. Shaji, A. Varol, P. Fua, Yashoteja, A. Jain, and S. Chandran. Resolving occlusion in multiframe reconstruction of deformable surfaces. In *Computer Vision and Pattern Recognition Workshops, 2011 IEEE Conference on*, pages 31–36, June 2011. 2

[23] S. Shen, W. Shi, and Y. Liu. Monocular 3-d tracking of inextensible deformable surfaces under l2-norm. *Trans. Img. Proc.*, 19(2):512–521, Feb. 2010. 2

[24] B. Zeisl, P. F. Georgel, F. Schweiger, E. Steinbach, and N. Navab. Estimation of location uncertainty for scale invariant feature points. In *Proceedings of the British Machine Vision Conference*, pages 57.1–57.12. BMVA Press, 2009. 4